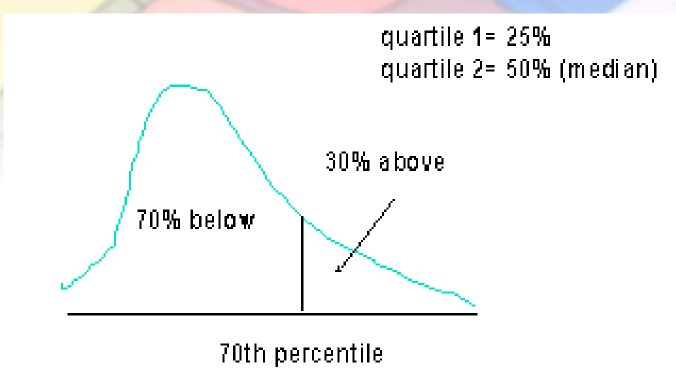
چندکها و
توزیع داده ها

مدرس : اصغر کاربرد اپیدمیولوژیست وعضو هیات علمی دانشگاه علوم پزشکی قزوین

karbord2003@yahoo.com

Percentiles



چنــدک ها :

عدد Q_pرا که در آن $0< P<1$ ، چندک مرتبهP می نامند ،

هرگاه $100P$درصد داده‌ها کوچکتر ازآن باشند ، مثلاً $Q_{/35}$ ،

چندک مرتبه ۳۵/. می‌باشد ، هرگاه تقریباً ۳۵ درصد داده‌ها

کوچکتر از$Q_{/35}$ باشد.

چندک‌های معروف عبارتند از :

الف)چارک ها

چارک ها که به ازای ۷۵/. و ۵/. و ۲۵/. بدست می‌آیند وآنها رابا Q_1وQ_2وQ_3 نشان می‌دهند که در آن Q_2 همان میانه می‌باشد.

ب)دهک‌ها

دهک‌ها که به ازای$P=/1 , /2 , ... , /9$ بدست می‌آیند وآن‌ها را باD_1و D_2و ... وD_9 نشان می دهند.

ج)صدک‌ها

صدک‌ها که بازای $P=/1 , /2 , ... , /9$ بدست می‌آیند و آنها را با P_1 و P_2و ... وP_9 نشان می‌دهند.

محاسبهٔ چندک‌ها در داده های طبقه بندی نشده

فرض کنید n داده به صورت غیر نزولی $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ مرتب شده باشد برای بدست آوردن Q_p ابتدا $(n+1)P$ را تشکیل می‌دهیم اگر حاصل ،عددی صحیح مانند k شود بنابراین $Q=X_{(k)}$ در غیر اینصورت: $(n+p)=k+r$

که در آن $۰<r<1$ وقسمت اعشاری $(n+1)P$ می‌باشد در این صورت با یک تناسب فاصله ای ساده داریم:

$$Q_p= X_{(k)}+r\,(X_{(k+1)}-X_{(k)})$$

مثال:درداده‌ها زیر ، چارک اول را محاسبه کنید.

۱۳و۱۲و۹و۸و۱۴و۱۶و۱۶و۱۷و۱۴و۱۳و۲۰و۱۶و۱۵و۱۱و۱۰و۷و۱۴و۶و۱۲

ابتدا داده ها را به صورت غیر نزولی مرتب می کنیم

۲۰و۱۷و۱۶و۱۶و۱۶و۱۵و۱۴و۱۴و۱۴و۱۳و۱۳و۱۲و۱۲و۱۱و۱۰و۹و۸و۷و۶

در این صورت چارک اول که همان صدک ۲۵م می باشد عبارت است از: $Q_1 = Q_{./25}$

$(n+1)P = 20\times ./25 = 4$ $\qquad$ $P = ./25$ n=۱۹

بنابراین $k=5$

و... $\qquad$ $Q_{./25} = X_{(5)} = 10$

انواع چارک ها

Q_1 : مقداری که ۲۵٪ مشاهدات ، پایین تر از آن است

Q_2 : مقداری که ۵۰٪ مشاهدات ، پایین تر از آن است

Q_3 : مقداری که ۷۵٪ مشاهدات ، پایین تر از آن است

karbord2003@yahoo.com

نحوه محاسبه چارکها

۱- مرتب نمودن صعودی داده ها

۲- کد گذاری کردن آنها از ۱ تا N

۳- پیدا نمودن محل چارک مورد نظر

۴- تعیین نمودن مقدار چارک مورد نظر به کمک محل چارک

karbord2003@yahoo.com

چارکهای نمونه

چارک (اول) کوچکتر	صدک ۲۵ ام=Q_1
چارک دوم (یا میانه)	صدک ۵۰ ام=Q_2
چارک (سوم) بزرگتر	صدک ۷۵ ام=Q_3

karbord2003@yahoo.com

مثال: ۹ کارگر صنعتی تحت آزمون قدرت پنجه قرار گرفتند و اندازه‌های زیر به دست آمدند:

۱۲۸/۳ ۱۰۶/۵ ۱۱۶/۶ ۱۵۲/۴ ۱۲۵/۰ ۱۳۲/۱ ۱۰۵/۸ ۱۳۶/۷

چارک‌ها را برای این داده‌ها محاسبه کنید.

karbord2003@yahoo.com

ابتدا داده‌ها را بترتیب از کوچکترین به بزرگترین اندازه مرتب می‌کنیم:

۱۵۲/۴ ۱۳۶/۷ ۱۳۲/۱ ۱۲۸/۳ ۱۲۵/۰ ۱۱۶/۶ ۱۰۶/۵ ۱۰۵/۸ ۹۳/۹

چارک اول:

تعدادمقادیر نابیشتر از Q_1 حداقل برابر است با ۰/۲۵×۹=۲/۲۵

تعدادمقادیر ناکمتر از Q_1 حداقل‌برابراست با ۰/۷۵×۹=۶/۷۵

مشاهده ۱۰۶/۵ دارای چنین شرایطی است و چارک اول است.

karbord2003@yahoo.com

۱۵۲/۴ ۱۳۶/۷ ۱۳۲/۱ ۱۲۸/۳ ۱۲۵/۰ ۱۱۶/۶ ۱۰۶/۵ ۱۰۵/۸ ۹۳/۹

با شمارش سه مشاهده از بزرگترین مقدار به طرف پایین، ۱۳۲/۱ چارک سوم است.

۱۵۲/۴ ۱۳۶/۷ ۱۳۲/۱ ۱۲۸/۳ ۱۲۵/۰ ۱۱۶/۶ ۱۰۶/۵ ۱۰۵/۸ ۹۳/۹

میانه یا صدک ۵۰ ام مقدار وسطی یعنی ۱۲۵/۰ است.

karbord2003@yahoo.com

توجه: معمولاً چارک‌ها را در حالاتی که تعداد مشاهدات **کمتر از ۲۵** است، محاسبه نمی‌کنند.

چارک اول – چارک سوم= دامنه میان چارکی نمونه

karbord2003@yahoo.com

محاسبۀچندک ها در داده های طبقه بندی شده

روش محاسبه چندک ها کاملاً شبیه محاسبه میانه است.برای محاسبه چندک Q_p ،ابتدا عدد n.p را با ستـــون فراوانی تجمـــعی مقایسه می‌کنیم اولین طبقه‌ای که فراوانی تجمعی آن بزرگتر یا مساوی n.p باشد راطبقۀچندکpام نامیده و با استفاده از یک تناسب ساده ریاضی Q_p را از فرمول زیر محاسبه می کنند.

$$Q_p = L_p + \frac{n.p - F_{(p-1)}}{f_p}.w$$

نمودارهاي آماري:

نما

چندکها

میانه

میانگین

محاسبه چندك براي داده‌هاي پيوسته

با توجه به ستون فراواني تجمعي در جدول فراواني، كلاسي را كه چندك در آن قرار دارد مشهص مي‌كنيم.

كلاس	x_i	f_i	F_i
۳۵/۱_۵۵/۱	۴۵/۱	۴	۴
۵۵/۱_۷۵/۱	۶۵/۱	۶	۱۰
۷۵/۱_۹۵/۱	۸۵/۱	۱۲	۲۲
۹۵/۱_۱۵/۲	۰۵/۲	۹	۳۱
۱۵/۲_۳۵/۲	۲۵/۲	۸	۳۹
۳۵/۲_۵۵/۲	۲/۴۵	۶	۴۵
۵۵/۲_۷۵/۲	۶۵/۲	۲	۴۷
۷۵/۲_۹۵/۲	۸۵/۲	۳	۵۰
جمع		۵۰	_

$$\underset{(p)}{0.25} \times \underset{(n)}{50} = 12.5$$

$$Q_p = L_{Q_p} + \left(\frac{np - F_b}{f_{Q_p}}\right)w$$

$$Q_p = 1.75 + \left(\frac{50 \times 0.25 - 10}{12}\right) \times 0.2$$

که در آن

L_p : حد پایین طبقهٔ چندک

f_p : فراوانی طبقهٔ چندک

F_{p-1}: فراوانی تجمعی طبقهٔ ماقبل چندک

W : طول طبقه

مثال:در جدول توزیع فراوانی سن ازدواج چارک اول و صدک ۶۰ام عبارتنداز:

سن ازدواج	فراوانی سن	فراوانی تجمعی
$14/5-19/5$	18	18
$19/5-24/5$	74	92
$24/5-29/5$	62	154
$29/5-34/5$	26	180
$34/5-39/5$	20	200

$n = 200$

$$Q_1 = Q_{0/25} = 19/5 + \frac{200\times 0/25 - 18}{74} \times 5 = 21/66$$

$$Q_{60} = Q_{./60} = 24/5 + \frac{20\times ./6 - 92}{62} \times 5 = 26/76$$

مثال: میانه و دامنه میان چاکی را برای دادهای زیر بدست آورید:

رده (سن برحسب سال)	فراوانی	فراوانی تجمعی
۱۴/۵-۱۹/۵	۱۸	۱۸
۱۹/۵-۲۴/۵	۷۴	۹۲
۲۴/۵-۲۹/۵	۶۲	۱۵۴
۲۹/۵-۳۴/۵	۲۶	۱۸۰
۳۴/۵-۳۹/۵	۲۰	۲۰۰
مجموع	۲۰۰	

برای محاسبه میانه،

p=۰/۵ np=۲۰۰(۰/۵)=۱۰۰

$$Q_p = L_p + \frac{n.p - F_{(p-1)}}{f_p}.w$$

$\boxed{\text{ردهی میانه } ۲۴/۵ - ۲۹/۵}$ فراوانی آن f=۶۲

میانه $= ۲۴/۵ + \frac{(۱۰۰-۹۲)}{۶۲} \times ۵ = ۲۵/۱۵$ سال

$Q_1 = ۱۹/۵ + \frac{۳۲}{۷۴} \times ۵ = ۲۱/۶۶$ سال

$Q_3 = ۲۴/۵ + \frac{(۱۵۰-۹۲)}{۶۲} \times ۵ = ۲۹/۱۸$ سال

$Q_3 - Q_1 = ۲۹/۱۸ - ۲۱/۶۶ = ۷/۵۲$ سال

رده (سن برحسب سال)	فراوانی	فراوانی تجمعی
۱۴/۵-۱۹/۵	۱۸	۱۸
۱۹/۵-۲۴/۵	۷۴	۹۲
۲۴/۵-۲۹/۵	۶۲	۱۵۴
۲۹/۵-۳۴/۵	۲۶	۱۸۰
۳۴/۵-۳۹/۵	۲۰	۲۰۰
مجموع	۲۰۰	

و) دامنۀ میان چارکی :

در توزیع هایی که دارای تعداد اندکی داده نامعلوم یا پرت ، یعنی خیلی کوچک یا خیلی بزرگ در ابتدا و انتهای منحنی فراوانی هستند از انحراف چارکی بعنوان شاخص پراکندگی استفـــــاده می شود.

$$IQR=Q_3-Q_1$$

گاهی اوقات به جای دامنۀ میان چارکی از نیمه میان چارکی با عنوان انحراف چارکی استفاده می شود که فرمول آن عبارت است از:

$$SIQR=\frac{Q_3-Q_1}{2}$$

مثال:برای داده های ۱۲و۸و۷و۷و۵و۴و۳و۲و۱و۰داریم:

$$IQR = Q_3 - Q_1 = 7 - 2 = 5$$

$$SIQR = \frac{Q_3-Q_1}{2} = \frac{7-2}{2} = 2/5$$

$$IQ = \frac{Q_1+Q_3}{2} = \frac{7+2}{2} = 4/5$$

ضریب تغییرات چارکی :

درتوزیعهایی که دارای اندکی دادهٔ نامعـــلوم یا پرت در ابتدا یا انتهای منحنی فراوانی هستند از معیار ضریب تغییرات چارکی استفاده می شودکه فرمول آن عبارت است از:

$$C.V.Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

مثال:توزیع فراوانی سن افراد جویای کار در جدول زیر آمده است.

	فراوانی	فراوانی تجمعی
کمتر از۱۲سال	۲	۲
۱۲-۱۶	۱۸	۲۰
۱۶-۲۰	۴۰	۶۰
۲۰-۳۰	۶۰	۱۲۰
۳۰-۴۰	۵۰	۱۷۰
۴۰-۶۰	۲۰	۱۹۰
بیش از ۶۰سال	۱۰	۲۰۰

$Q_1 = 19$

$Q_3 = 36$

در این صورت

$$C.V.Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{36-19}{36+19} = \frac{17}{55} = 0/31$$

توزیع نرمال

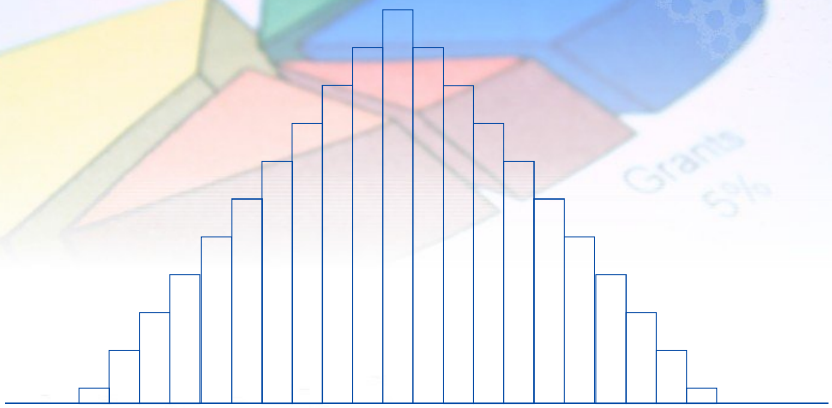
یعنی استاندارد کردن متغیر $X \approx \mathcal{N}(\mu , \delta)$
با استفاده از متغیر نرمالی که میانگین آن
صفر و واریانس اش یک است و سپس
استفاده از جدول مربوطه

karbord2003@yahoo.com

منحنی طبیعی به وسیله گاووس ریاضی دان آلمانی کشف شده است . گاووس توزیع خطاهایی را که در مشاهده های ستاره شناسی انجام شده بود مطالعه کرد و آنها را به صورت نمودار نشان داد. به همین خاطر منحنی طبیعی را منحنی گاووس نیز می گویند.

نکته: منحنی خیلی از حوادث فیزیکی و روانشناسی مانند قد، وزن، طبیعی است.

karbord2003@yahoo.com



هیستوگرام متقارن

مفهوم توزیع طبیعی در آمار یک مفهوم اساسی و بنیـــادی است. دلیل اهمیت آن این است که توزیع فراوانی هیستوگرام بسیاری از حوادث طبیعی شکل منحنی طبیـــعی یا نرمال را دارند یعنی فراوانی آنها ابتدا به یک مقدار بیشـــینه صعود می‌کند و سپس به طور متقارن کاهش می‌یابد.

با آنکه بسته به مقدار میانگین و انحراف معیار تعداد بی نهایت منحنی نرمال وجود دارد ولی تنها یک منحنی نرمال استاندارد توسط امارشناسان به منظور سهولت برآورد مساحت منحنی نرمال در نظر گرفته می شود. منحنی نرمال استاندارد یک منحنی هموار ، زنگوله ای شکل و کاملا قرینه بر اساس تعداد نامحدودی نشانه است. مجموع مساحت زیر منحنی یک، میانگین صفر و انحراف معیار آن یک است. میانگین میانه و نما در منحنی نرمال استاندارد بر هم منطبق می باشند و فاصله هر مقدار (X) از میانگین منحنی بر حسب واحد انحراف معیار به نامه انحراف نسبی یا تغییر نرمال استاندارد نامیده و بطور معمول با Z نشان داده می شود. مقدار Z از رابطه زیر بدست می آید:

$$Z = \frac{X - \mu}{\sigma}$$

ویژگیهای منحنی طبیعی

۱- منحنی طبیعی متقارن است و حداکثر ارتفاع آن در میانگین قرار دارد.

۲- در منحنی طبیعی میانگین، میانه و نما بر روی هم قرار دارند.

۳- منحنی دو نقطه عطف دارد که شکل منحنی در آن دو نقطه تغییر می کند.

۴- دنباله های منحنی با محور X موازی هستند. بنابراین منحنی از $+\infty$ تا $-\infty$ ادامه دارد ولی در عمل این منحنی از ۳- تا ۳+ می باشد.

۵- منحنی طبیعی به شکل زنگوله است.

نکته: اختلاف در شکل منحنی های طبیعی به دلیل اختلاف در پراکندگی آنها است.

منحنی طبیعی استاندارد

برای یکسان کردن منحنیهای توزیعهای مختلف طبیعی می توان آنها را به توزیع نمره های استاندارد (نمره Z)تبدیل کرد.

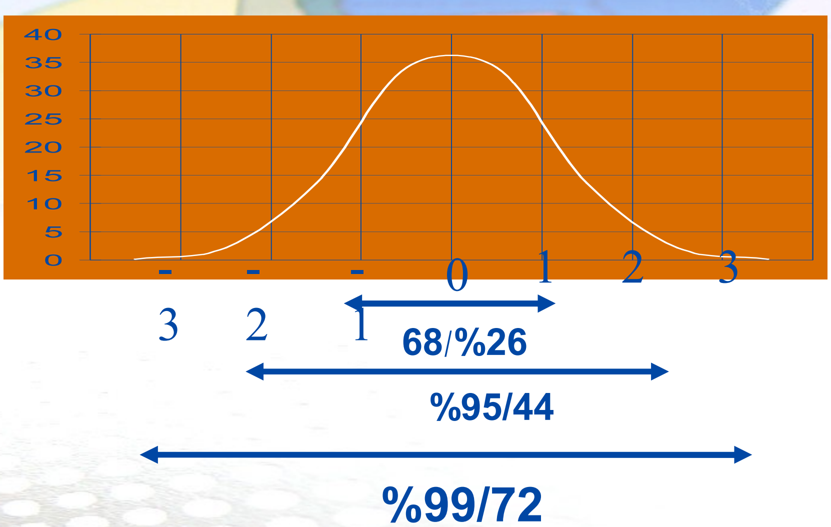
بنابراین هنگامی که نمره های توزیع به نمره های استاندارد تبدیل می شود توزیعی با میانگین صفر و انحراف استاندارد یک بدست می آید که به آن منحنی طبیعی استاندارد می گویند.

$$Z = \frac{X - \overline{X}}{S}$$

نکته: محور(X) منحنی طبیعی استاندارد براساس Z تقسیم بندی می شود نه نمره خام .

نکته: بهترین مرز در منحنی استاندارد میانگین می باشد که Z آن برابر با صفر است که در واقع همان میانه و نما هم می باشد.

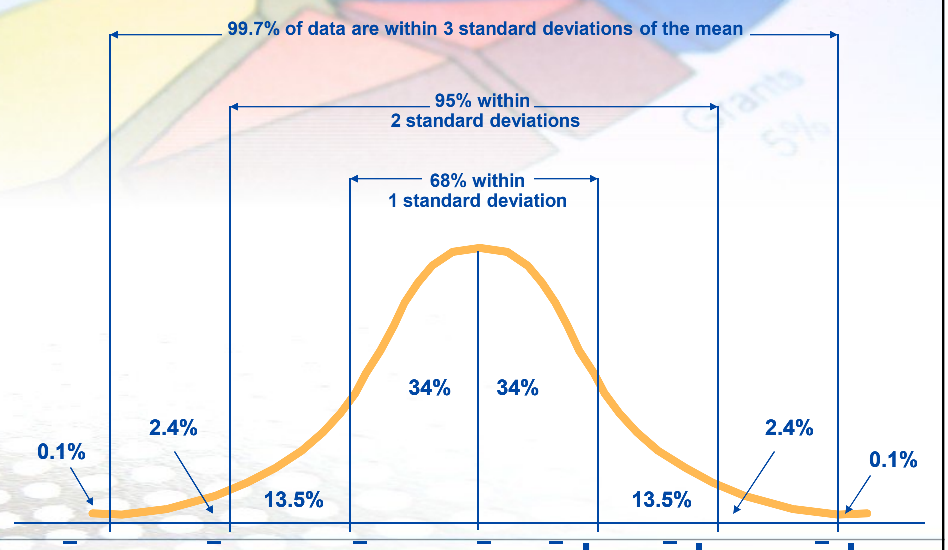
سطح زیر منحنی طبیعی استاندارد



درصد های توزیع نرمال

- در هر توزیع نرمال ٪۶۸ از داده ها در فاصله ۱ انحراف معیار از میانگین قرار دارند.
- در هر توزیع نرمال ٪۹۵ از داده ها در فاصله ۲ انحراف معیار از میانگین قرار دارند.
- در هر توزیع نرمال ٪۹۹/۷ از داده ها در فاصله ۳ انحراف معیار از میانگین قرار دارند.

خواص توزیع نرمال
Normal Distribution



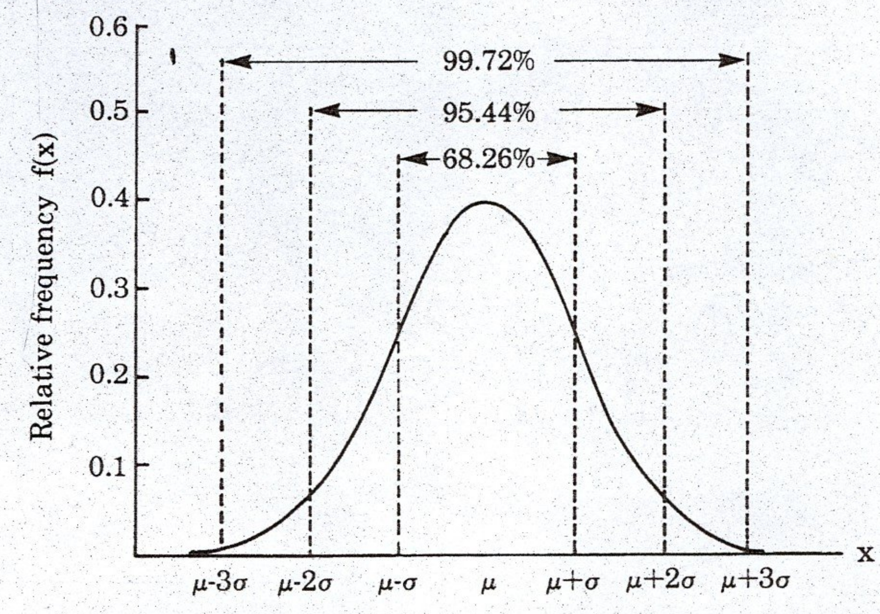
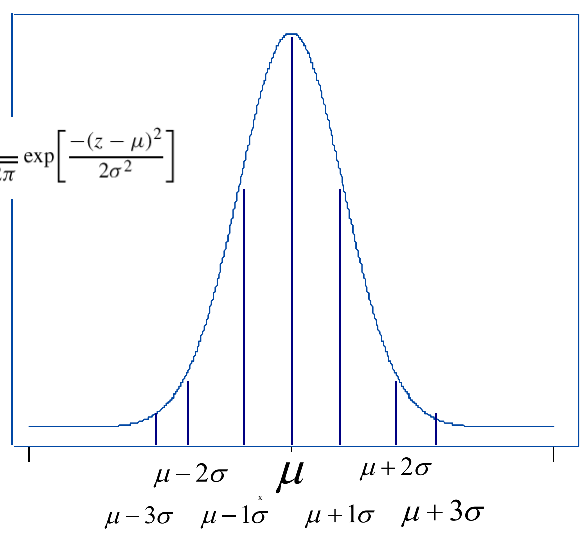


FIGURE 11-5. The Gaussian probability distribution.

Areas Under the Normal Curve

$$f(z|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(z-\mu)^2}{2\sigma^2}\right]$$



Between:
68.26%
95.44%
99.74%

توزیع نرمال

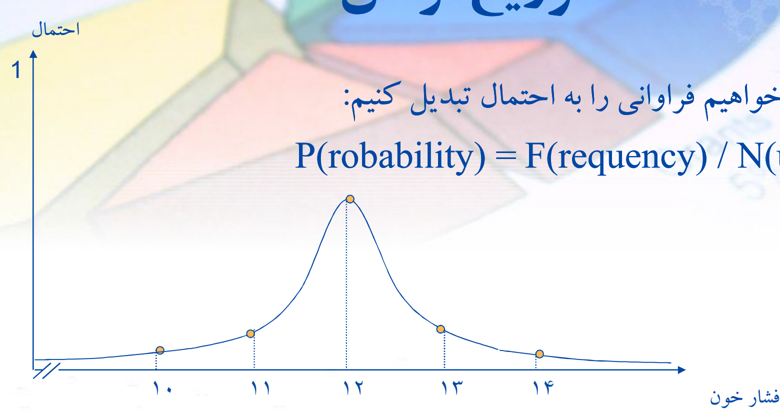
- ✓ میانگین فشار خون جامعه
- ✓ سرشماری فشار خون در آمریکا: وضعیت چندان تفاوتی نکرده
- ✓ فراوانی فشار خون ها (اعداد طبیعی)
- ✓ فشار خون کمیتی پیوسته است

karbord2003@yahoo.com

توزیع نرمال

- ✓ می خواهیم فراوانی را به احتمال تبدیل کنیم:

P(robability) = F(requency) / N(umber)



جمع تمام این احتمالات چند است؟ (سطح زیر منحنی چقدر است؟)

این منحنی چه چیز را نشان می دهد؟ (چگونگی توزیع احتمال فشار خون های مختلف در جامعه)

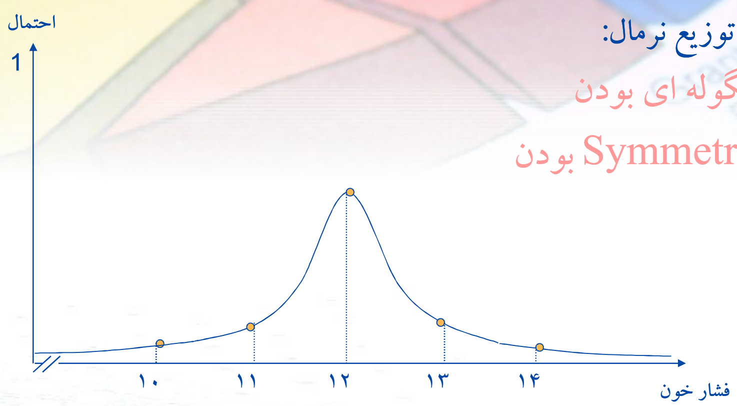
به این طور نمودارها، نمودارهای نرمال (توزیع نرمال) می گویند

karbord2003@yahoo.com

توزیع نرمال

خصوصیات توزیع نرمال:

- ✓ زنگوله ای بودن
- ✓ Symmetric بودن



karbord2003@yahoo.com

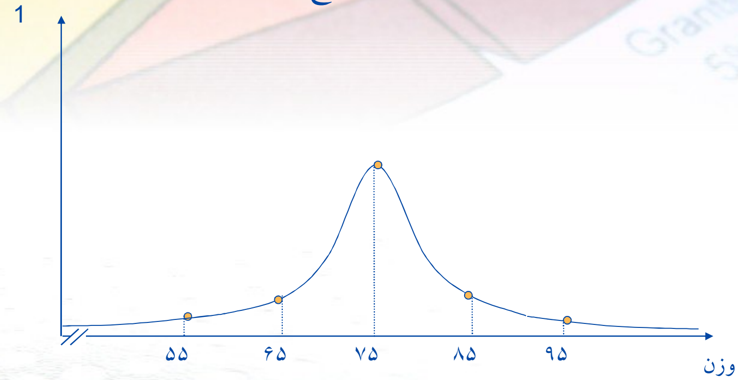
توزیع نرمال

شاخص های نماینده توزیع نرمال:

karbord2003@yahoo.com

توزیع نرمال

شاخص های نماینده توزیع نرمال:

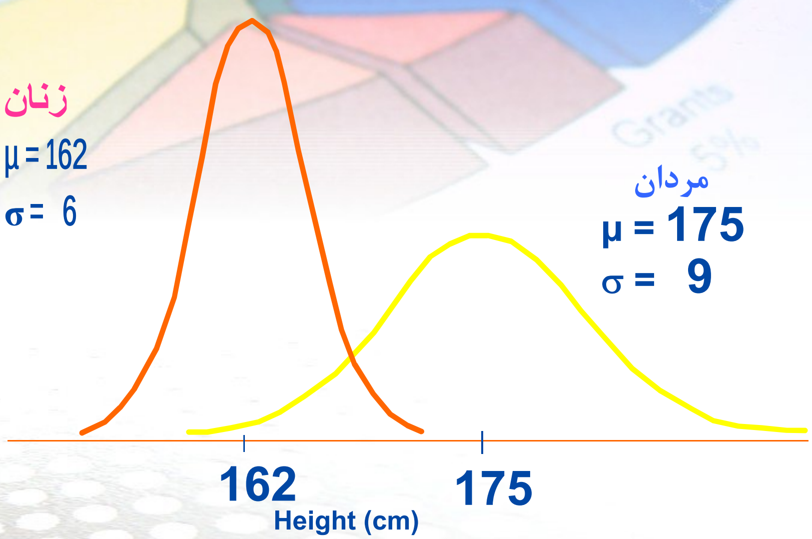


karbord2003@yahoo.com

توزیع قد مردان و زنان

- زنان
- $\mu = 162$
- $\sigma = 6$

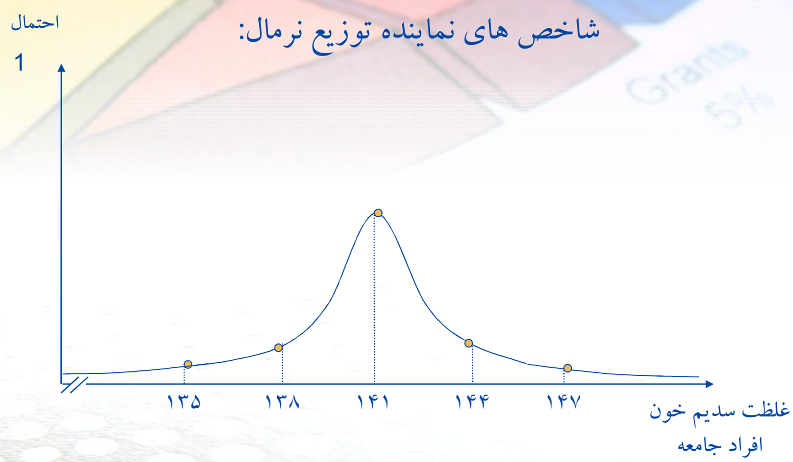
مردان
$\mu = 175$
$\sigma = 9$



karbord2003@yahoo.com

توزیع نرمال

شاخص های نماینده توزیع نرمال:

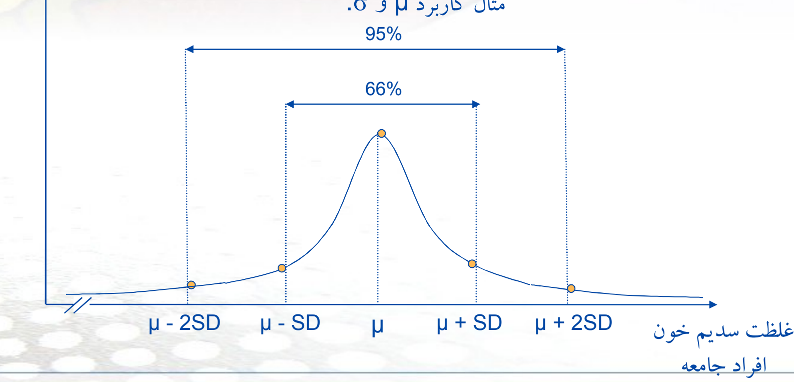


توزیع نرمال

شاخص های نماینده توزیع نرمال:

۱. میانگین (μ): نشان دهنده تراکم داده ها حول یک مقدار
۲. انحراف معیار (σ): نشان دهنده میزان پراکندگی داده ها از میانگین

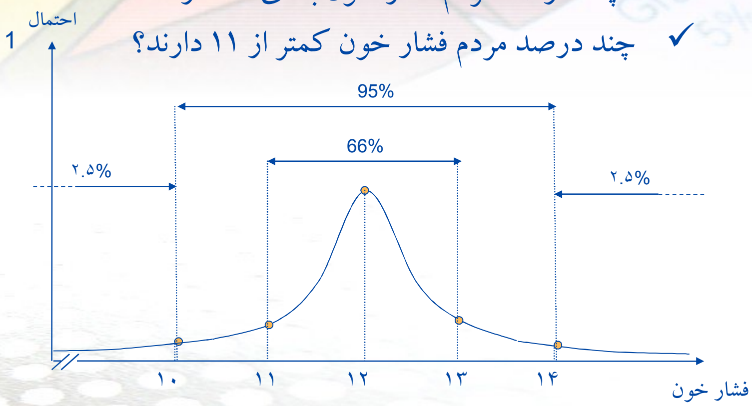
مثال کاربرد μ و σ:



توزیع نرمال

تمرین ۱: اگر میانگین فشار خون در جامعه ۱۲ و انحراف معیار آن ۱ باشد:

- ✓ چند درصد مردم فشار خون بالای ۱۴ دارند؟
- ✓ چند درصد مردم فشار خون کمتر از ۱۱ دارند؟



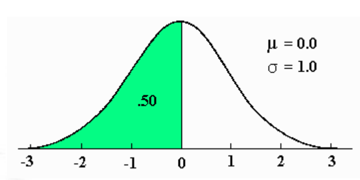
توزیع نرمال استاندارد

اگر برای متغیر X با توزیع نرمال تغییر متغیر

$$Z = \frac{X-\mu}{\sigma}$$

را در نظر بگیریم آنگاه متغیر Z دارای توزیع نرمال استاندارد خواهد بود که میانگین آن صفر و انحراف معیار آن یک می باشد.

$$Z \sim N(0,1)$$

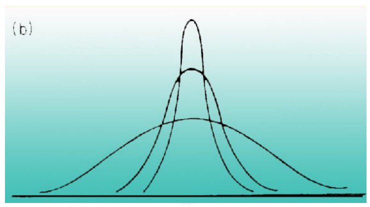


در جدول نرمال استاندارد سطح زیر منحنی از نقطه صفر تا نقطه Z محاسبه شده است .
برای محاسبه سطح زیر منحنی در قسمت منفی از خاصیت تقارن استفاده می شود.

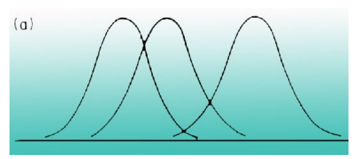
پارامتر های توزیع نرمال

۱- میانگین μ $\quad -\infty < \mu < +\infty$

۲- انحراف معیار σ $\quad 0 < \sigma < +\infty$



تغییرات انحراف معیار



تغییرات میانگین

منحنی طبیعی دارای معادله ای به شرح زیر است که ارتفاع منحنی را در هر نقطه از X نشان می دهد.

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

میانگین = μ

Y = ارتفاع منحنی در هر نقطه ای از X

X = متغیر اندازه گیری شده

σ = انحراف استاندارد

محاسبه احتمالات توزیع نرمال

- روش تبدیل هر متغیر نرمال(X) به نرمال استاندارد(Z):

$$Z = \frac{X - \mu}{\sigma}$$

- **مثال:**
- اگر مقادیر فشار خون سیستولیک افراد سالم جامعه دارای توزیع نرمال با میانگین ۱۲۰ و انحراف معیار ۱۰ میلیمتر جیوه باشد:
- ۱- فشار خون چند درصد افراد بیش از ۱۳۰ میلیمتر جیوه است؟

X=130 Z=(130-120)/10=1.00 P=0.159=15.9%

کاربردهای توزیع میانگین نمونه

- قبلا دیدیم برای تبدیل X به Z از رابطه زیر استفاده می شد:

$$Z = \frac{X - \mu}{\sigma}$$

- در اینجا نیز رابطه تبدیل $\bar{X}$ به Z به صورت زیر است:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

- مثال ها :
- ۱- در ادامه مثال گذشته، فرض کنیم در یک نمونه ۲۵ نفری از زنان و مردان بین ۲۰ تا ۳۴ ساله، میانگین فشار خون سیستولیک آنان ۱۲۴ میلیمتر جیوه به دست آمده است. چقدر ممکن است در نمونه ۲۵ نفری، میانگین نمونه ۱۲۴ یا بیشتر از آن باشد؟
- Z=(124-120)10/√25=4/2=2.00 P=0.023=2.3%

انواع حالات توزیع ها

۱- متقارن (نرمال) : مد = میانه = میانگین

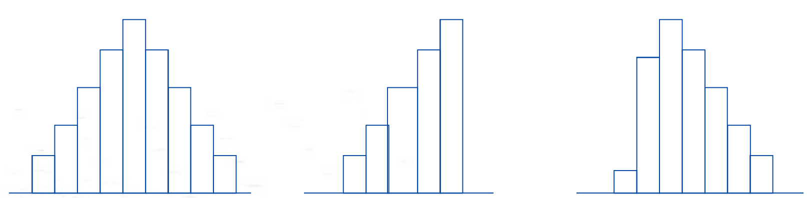
۲- چوله به راست : مد > میانه > میانگین

۳- چوله به چپ : مد < میانه < میانگین

karbord2003@yahoo.com

چولگی:

میزان عدم تقارن منحنی را چولگی می‌گویند. در توزیعهای چوله معیارهای گرایش مرکزی بر هم منطبق نیستند ولی همیشه میانه بین نما و میانگین قرار دارد. اگر نما از میانگین بزرگتر باشد منحنی چوله به چپ است یعنی دم بلند منحنی به طرف چپ می‌باشد و اگر میانه از میانگین کوچکتر باشد منحنی چوله به راست می‌باشد .

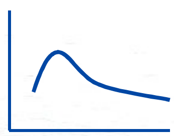


هیستوگرام چوله به راست — هیستوگرام چوله به **چپ** — هیستوگرام متقارن

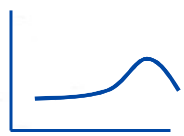
karbord2003@yahoo.com

مفهوم چولگی

اگر دم توزیع جامعه به سمت راست باشد ، توزیع را چوله به راست و در صورت عکس ، آن را چوله به چپ می نامند



چوله به راست



چوله به چپ

karbord2003@yahoo.com

مقادیر مختلف ضریب چولگی SK

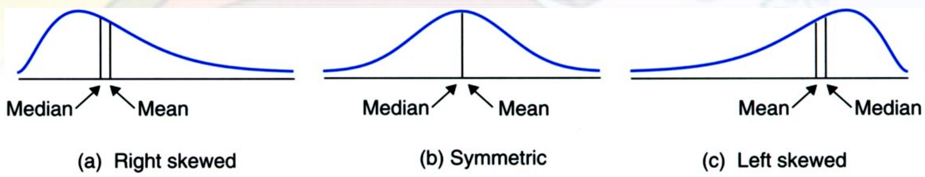
۱- صفر : در صورت متقارن بودن توزیع جامعه

۲- مثبت : در صورت چوله به راست بودن توزیع جامعه

۳- منفی : در صورت چوله به چپ بودن توزیع جامعه

karbord2003@yahoo.com

Relative positions of the mean and median for (a) right-skewed, (b) symmetric, and (c) left-skewed distributions

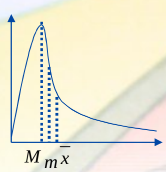


Note: The mean assumes that the data is normally distributed. If this is not the case it is better to report the median as the measure of location.

منحنيهاي فراواني:

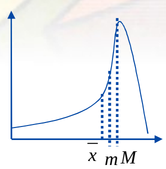
چولگي:

به راست يا مثبت



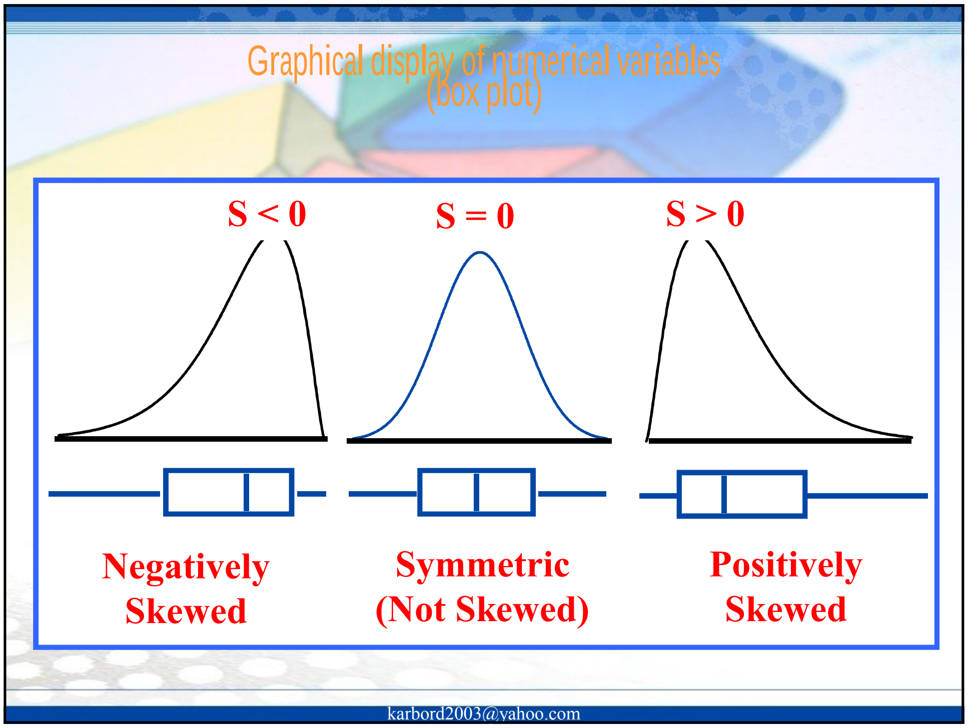
$M < m < \bar{x}$

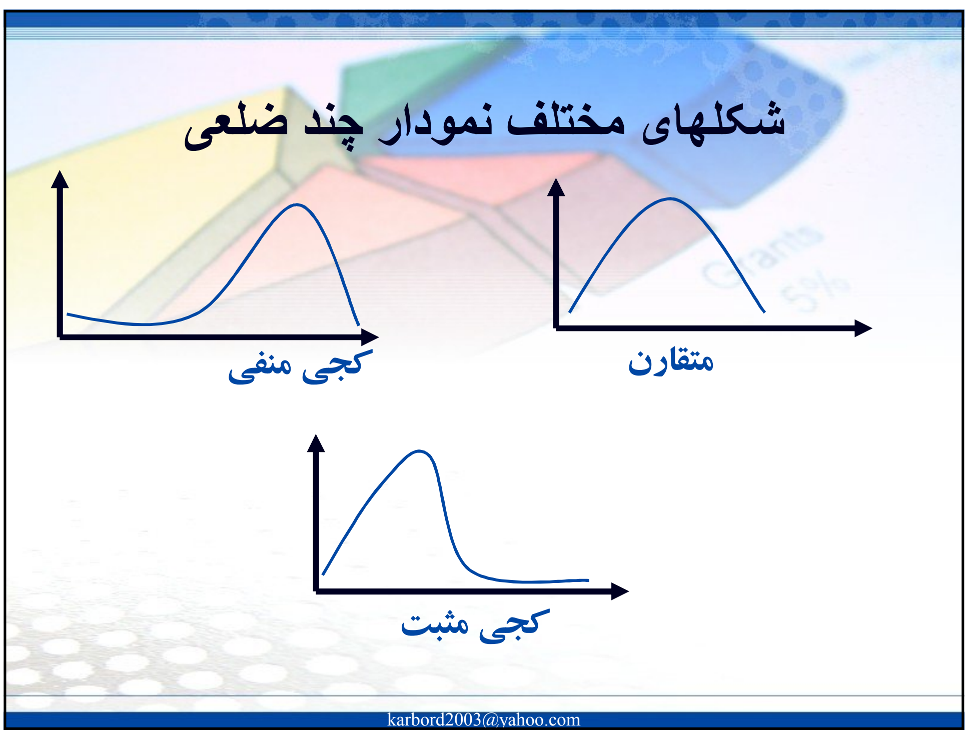
به چپ يا منفي



$M > m > \bar{x}$

معیار اندازه گیری چولگی: $\frac{\bar{x}-M}{s} =$ ضريب چولگی اول پيرسن


Graphical display of numerical variables
(box plot)
S < 0
S = 0
S > 0
Negatively
Skewed
Symmetric
(Not Skewed)
Positively
Skewed
karbord2003@yahoo.com


شکلهای مختلف نمودار چند ضلعی
کجی منفی
متقارن
کجی مثبت
karbord2003@yahoo.com

تفاوت میانگین و نما $(\bar{x}-M)$ یا تفاوت آن با میانه $(\bar{x}-m)$ می‌تواند مشخص کننده چولگی باشد حال این اختلافها با تقسیم بریک معیار پراکندگی هم‌بعد مانند S از قید واحد خارج می‌شوند .
بااین انگیزه پیرسون برای اندازه گیری میزان و جهت چولگی ضرائب چولگی زیر را تعریف نمود.

ضریب اول چولگی پیرسون

$$sk_1 = \frac{\bar{x} - M}{s}$$

$$sk_2 = \frac{3(\bar{x} - m)}{s}$$

$$sk_3 = \frac{m_3}{s^3}$$

در توزیعهایی که چولگی زیاد نباشد رابطه تجربی $\bar{x} - m \approx 3(\bar{x} - M)$ که به رابطه پیرسون شهرت دارد برقرار است.

تفسیر ضرایب چولگی

اگر sk=۰ باشد توزیع چولگی ندارد.

اگر sk > ۰ باشد توزیع چوله به راست است .

اگر sk < ۰ باشد توزیع چوله به چپ است .

در عمل ضرائب چولگی با اعداد ۱/۰ و ۵/۰ مقایسه می شوند بدین ترتیب که :

اگر |sk| < ۱/۰ باشد توزیع چوله نیست و می تواند متقارن باشد.

اگر ۵/۰ > |sk| > ۱/۰ باشد توزیع چوله است و هرچه sk به ۱/۰ نزدیک باشد چولگی کم و هر چه به ۵/۰ نزدیک باشد چولگی زیاد است .

اگر |sk| > ۵/۰ باشد توزیع خیلی چوله است.

karbord2003@yahoo.com

مثال: برای توزیع سن بیمارانی که به یک بیمارستان کودکان مراجعه کرده‌اند و به شرح زیر است.ضریب چولگی اول پیرسون را محاسبه و تفسیر کنید .

x_i	f_i	F_i	$f_i x_i$	$f_i x_i^2$
2	4	4	8	16
3	40	44	120	360
4	24	68	96	384
5	20	88	100	500
6	8	96	48	288
7	4	100	28	196

$n = 100 \quad \sum f_i x_i = 400 \quad \sum f_i x_i^2 = 1744$

karbord2003@yahoo.com

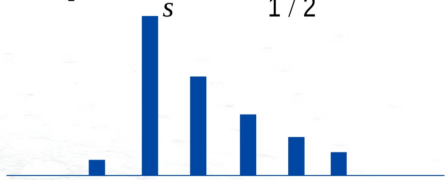
میانگین $= \bar{x} = \frac{\sum f_i x_i}{n} = \frac{400}{100} = 4$

نما $M = 3$

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \bar{x}^2}$$

$$s = \sqrt{\frac{1744}{100} - (4)^2} = 1/2$$

$sk_1 = \frac{\bar{x} - M}{s} = \frac{4-2}{1/2} = +/83$ بنابراین چولگی زیاد و مثبت است.



نمودار چوله به راست است

karbord2003@yahoo.com

مثال: طول عمر ۱۰۰ باطری اتومبیل دارای میانگین،میانه و انحراف استاندارد ۵/۳ و ۴۸/۳ و ۶۵/۱ سال می باشد ضریب چولگی با توجه به اطلاعات داده شده عبارت است از :

$\bar{x} = 2/5 \qquad m = 3/48 \qquad s = 1/65$

$$sk_1 = \frac{3(\bar{x} - m)}{s}$$

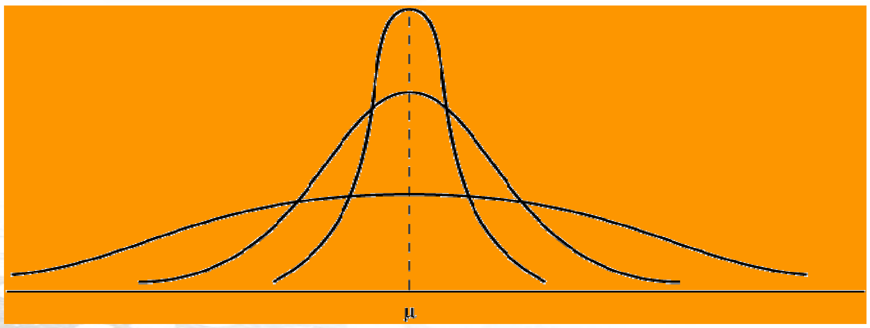
$$= \frac{3(3/5 - 3/48)}{1/65} = 0/036$$

karbord2003@yahoo.com

کشیدگی :

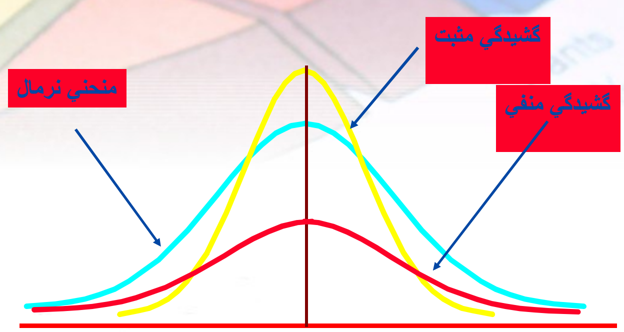
دو توزیع ممکن است دارای میانگین یکسان باشند و کاملاً نیز قرینه باشند ولی از لحاظ کشیدگی با هم تفاوت زیادی داشته باشند .

میزان برجستگی یا پخی منحنی فراوانی نسبت به منحنی مقایسه را کشیدگی گویند و منحنی مقایسه را محنحنی نرمال در نظر می گیرند.



karbord2003@yahoo.com

بررسی توزیع داده ها از لحاظ گشیدگی



karbord2003@yahoo.com

تفسیر ضریب کشیدگی :

توزیع نرمال است هرگاه k = ۰ باشد.

توزیع از توزیع نرمال کشیده‌تر است هرگاه k > ۰ باشد .

توزیع از توزیع نرمال پخ‌تر است هرگاه k<0 باشد.

توزیع تقریباً نرمال است هرگاه |k| < ۱/۰ توزیع یا توزیع نرمال تفاوت فاحش دارد هرگاه ۱/۰ < |k| < ۵/۰ باشد .

تفاوت توزیع با توزیع نرمال فاحش است هرگاه |k| > ۵/۰

karbord2003@yahoo.com

انواع توزیع به لحاظ کشیدگی و مقدار ضریب (E) آن

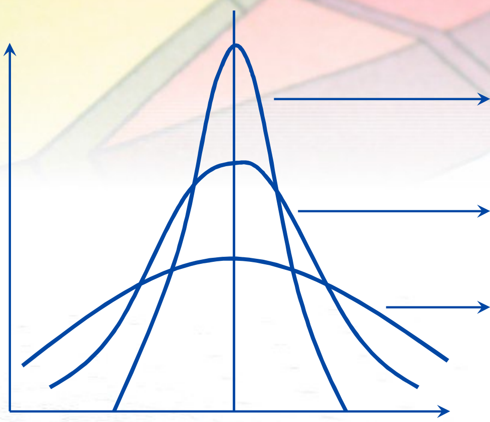
۱- مساوی توزیع نرمال (E=0)

۲- بلندتر از توزیع نرمال (E>0)

۳- کوتاه تر از توزیع نرمال (E<0)

karbord2003@yahoo.com

مقایسه انواع کشیدگی



$$\mu_x = Md = Mo$$

karbord2003@yahoo.com

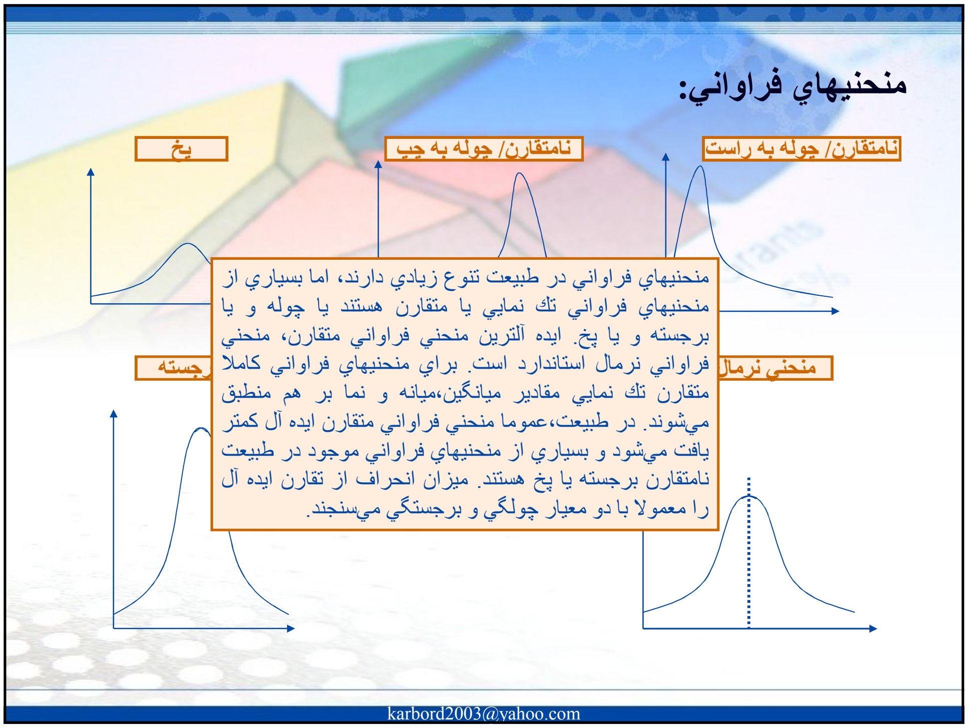
تفسیر ضریب کشیدگی (k)

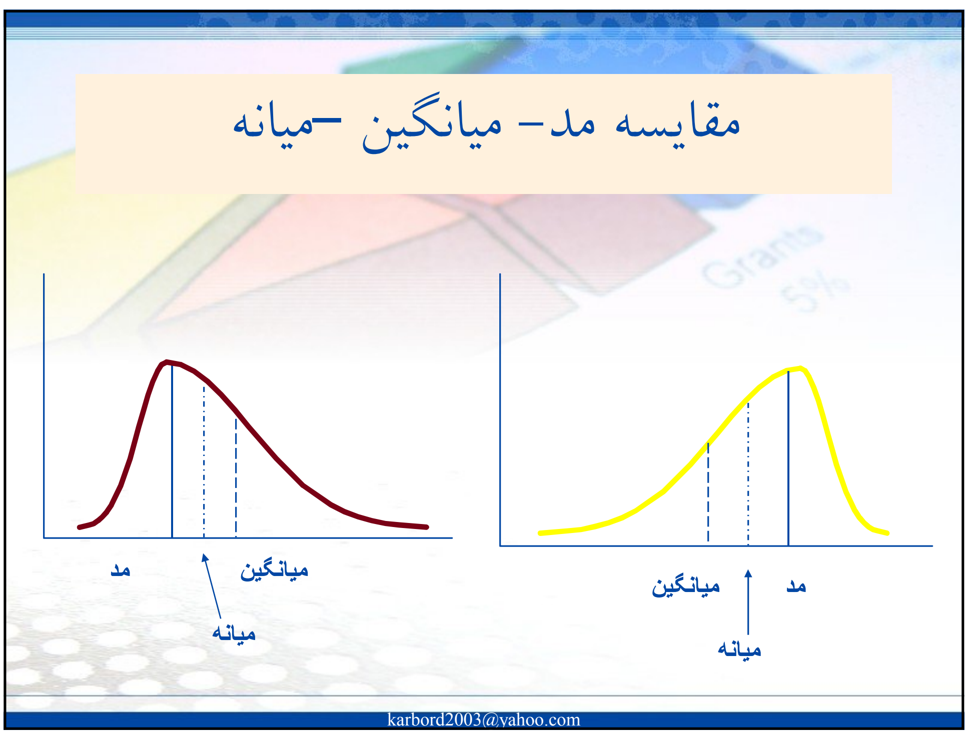
۱- توزیع نرمال $|E| \leq 0/1$

۲- توزیع نسبتاً بلندتر از نرمال $0/1 < |E| \leq 0/5$

۳- توزیع کاملاً کشیده تر از نرمال $|E| > 0/5$

karbord2003@yahoo.com


منحنيهاي فراواني:
پخ
نامتقارن/ چوله به چپ
نامتقارن/ چوله به راست
منحني نرمال
منحنيهاي فراواني در طبيعت تنوع زيادي دارند، اما بسياري از منحنيهاي فراواني تك نمايي يا متقارن هستند يا چوله و يا برجسته و يا پخ. ايده آلترين منحني فراواني متقارن، منحني فراواني نرمال استاندارد است. براي منحنيهاي فراواني كاملا متقارن تك نمايي مقادير ميانگين،ميانه و نما بر هم منطبق مي‌شوند. در طبيعت،عموما منحني فراواني متقارن ايده آل كمتر يافت مي‌شود و بسياري از منحنيهاي فراواني موجود در طبيعت نامتقارن برجسته يا پخ هستند. ميزان انحراف از تقارن ايده آل را معمولا با دو معيار چولگي و برجستگي مي‌سنجند.
karbord2003@yahoo.com


مقايسه مد- ميانگين –ميانه
مد
ميانگين
ميانه
ميانگين
مد
ميانه
karbord2003@yahoo.com

سوال

- برای داشتن اطمینان بیشتر، محدوده اطمینان چگونه تغییر می کند؟

$$\bar{X} \pm Z\sigma/\sqrt{N}$$

Confidence	$\alpha/2$	Z score
90%	0.05	1.65
95%	0.025	1.96
99%	0.005	2.58

karbord2003@yahoo.com

karbord2003@yahoo.com